Report Of

LOGO DETECTION

Done Under

RGSTC-TIFAC-MSME INTERNSHIP SCHEME

AT
KONVERGE.AI

Submitted By
Surbhi Khatri
Anirudha Agrawal
(VII , CSE)

Guide
Dr. Shailendra S. Aote

Department Of Computer Science and Engineering
Shri Ramdeobaba College of Engineering and Management
2019 - 20

ACKNOWLEDGEMENT

The written word has an unfortunate tendency to convert genuine appreciation into an embossed formality. But this is solitary mode by which we can record our approaches everlastingly. Numerous individuals have contributed their time and energies in aiding us complete our disserting labor magnificently.

We would like to express our sincere gratefulness to our guide Dr. Shailendra S. Aote , Computer Science and Engineering Department whose uninterrupted supervision and stimulation facilitated us to triumph grand altitudes. We express our gratitude to our company Konverge.ai pvt. Ltd. for their constant reinforcement and motivation.

We are very thankful to Dr. Rupesh Pais, Coordinator - RGSTC - TIFAC - MSME Program, Dr. M. B. Chandak Head, Computer Science and Engineering Department and Dr. R. S. Pande, Principal, RCOEM, for their constant support and help which encouraged us a lot in completing this internship. And last but not the least we would like to acknowledge our parents, friends, contemporaries and all those people known or unknown individuals who have directly or indirectly helped us in carrying out our work successfully.

Summary

Objective:

1. To provide Brand Detection as a Service.

2. To build a Deep learning tool to find brand logos in everyday pictures.

3. To build a general-purpose logo detection Software. To avoid re-training the network for each new company using the service, we divide logo detection and identification in two logically and operationally separate parts:

   I) To find all logos in the image with a YOLO detector (using the Keras implementation of keras-yolo3).

   II) To check for similarity between the proposed logos and an input uploaded by the customer.

4. To implement few shot learning on new logos without re-training the model again.

Results and Recommendation:

The logo Detection software has been developed and tested successfully on random images .It properly detected the possible logo containing regions and then correctly classified the detected logo. Thus it is working properly . Recommendation for the future scope of this unit is to train it on even larger data set to get further more accurate results and can use special zooming cameras which can be used to detect logos from a considerable distance too.We can also extend the problem statement to detect the if the logo in the image is original or fake.

TABLE OF CONTENTS

Chapter 1
Introduction

Konverge.Ai Private Limited is a Private incorporated on 21 March 2018. It is classified as Non-govt company and is registered at Registrar of Companies. It is involved in computer related activities Advanced Data Analytics , Computer Vision and Natural Language Processing.

They apply cutting-edge machine learning algorithms on the data to extract meaningful information for building an effective decision support system which can be used to make :

- Better predictions & forecasting
- Smart recommendation system
- Pattern recognition
- Anomaly & fraud detection

Their team of visionary engineers can provide customized end-to-end solutions to your organization using deep learning algorithms that focus on –

- Face Recognition
- Image classification
- Video analysis
- Object recognition
- Pattern identification

They deliver end-to-end solutions that integrate seamlessly with clients existing system and technology stack and converge on better, faster and efficient solutions using machine learning and artificial intelligence.

1.1 Challenges:

- Logo recognition is a challenging object recognition and classification problem as there is no clear definition of what constitutes a logo.

- A logo can be thought of as an artistic expression of a brand, it can be either a (stylized) letter or text, a graphical figure or any combination of these.

- Furthermore, some logos have a fixed set of colors with known fonts while others vary a lot in color and specialized unknown fonts.

- Due to the nature of a logo (as brand identity) there is no guarantee about its context or placement in an image, in reality logos could appear on any product, background or advertising surface.

- This problem has large intra-class variations e.g. for a specific brand, there exist various logos types (old and new Adidas logos, small and big versions of Nike) and inter-class variations e.g. there exists logos which belong to different brands but look similar (see Figure 1).

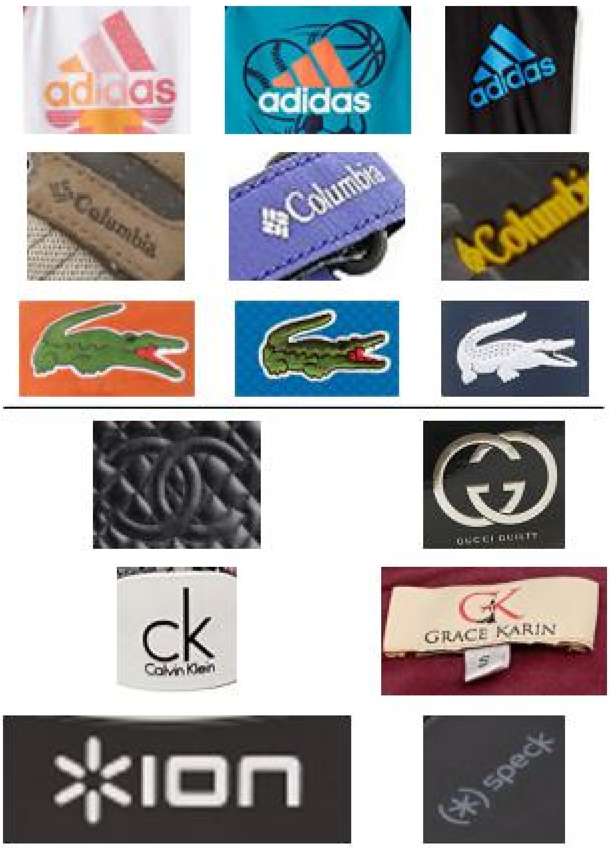


Figure 1. Logo variations exemplar images. (Row 1-3) Intra-class variations of brands Adidas, Columbia, Lacoste per row. Notice, different backgrounds, placement, fonts.

(Row 4-6) Inter-class variations of brands Chanel - Gucci, Calvin Klein - Grace Karin, Ion - Speck. Notice, similar looking logos but belong to different brands.

Problem Statement:

To detect and recognize logos from everyday images or videos using deep learning models to achieve better accuracy.

Research Objectives

- To develop a software for logo detection with highest accuracy.
- To gather dataset from various sources.
- To implement Universal logo Detector.
- To implement few shot learning to recognize new logos without training the model again

Scope and Limitations of the Research

The scope of this research can be extended to following points :

- Logo recognition to identify and monitor crises.
- Logo recognition for brand protection.
- Logo recognition for evaluating sponsorship.
- Logo recognition to find true brand association.

Limitations of the research are that logos can be of any shape , color , pattern , image of text. There is no fixed design of a logo and there are many intra-class and inter class variations which makes it difficult to always classify correctly .

Few-shot recognition becomes hard when image resolution is low and/or logos are made of very simple shapes.

Chapter 2
Literature Review

A. Watve and S. Sural (2008) [6], Soccer video processing for the detection of advertisement billboards, Billboards are placed on the sides of a soccer field for advertisement during match telecast. Unlike regular commercial, these are introduced while a break, on-field billboard appears on the TV screen at uncertain time instances, in different sizes, and also for different durations. L Ballan, m. Bertini and a. Jain (2008) [5], A system will automatically display the trademarks and technical broadcasting sporting events in US trademark evaluation of visibility of a sports marketing firm in cooperation with the aim of detecting our trademark and identity system that has been developed Show the current version of the game, videos, spread in order to identify. They propose a semiautomatic system and trademark appearances game, videos to return a human annotator trademarked time and an interface that shows the position of the detected through monitoring of the results of automatic annotation; So that the supervisor of your work fast to secure the parts of the video that a trademark has been marked as not containing can leave a good recall system for providing the data, the purpose of this is due to the fact.

L Ballan, m. Bertini, a. Del Bimbo, El Seidenari, and g. Serra (2011) [3], and video of the event, to explore the meaning of research on methods to represent recognition for annotations and recognition events and actions in the video, from the scientific community a growing attention to semantic indexing intelligent video surveillance systems and advanced human-computer interaction interface to video for many applications due to its relevance. In the proposed paper we survey the field of event recognition, from the interest point’s detector with descriptors, to event modelling techniques and knowledge management technologies. They provide an overview of the methods, categorising them according to video production methods and video domains, and according to type of actions or events that are typical of these domain.

Istvan Fehervari and Srikar Appalaraju (2018) Logo recognition is the task of identifying and classifying logos. Logo recognition is a challenging problem as there is no clear definition of a logo and there are huge variations of logos, brands and re-training to cover every variation is impractical. In this paper, we formulate logo recognition as a few-shot object detection problem. The two main components in our pipeline are universal logo detector and fewshot logo recognizer. The universal logo detector is a classagnostic deep object detector network which tries to learn the characteristics of what makes a logo. It predicts bounding boxes on likely logo regions. These logo regions are then classified by logo recognizer using nearest neighbor search, trained by triplet loss using proxies.

Chapter 3
Methodology

As discussed earlier, one of the key challenges with logo detection is that the context in which the logo is embedded can vary almost infinitely. State of the art deep learning object detectors that are trained to localize and identify a closed set of logos will inherently use the contexts of each logo for training and prediction which makes them susceptible to context changes. For example, a logo that appears only on shoes in the training data might remain invisible or get confused by same logo if it is displayed on a coffee mug during inference.

To avoid re-training the network for each new company using the service, logo detection and identification are split in two logically and operationally separate parts: first, we find all logos in the image with a YOLO detector (using the Keras implementation of keras-yolo3), and then we check for similarity between the proposed logos and an input uploaded by the customer (for example, the company owning the logo), by computing cosine similarity between features extracted by a pre-trained Inception network.

3.1 LOGO DETECTION

Given a large number of training images across a wide range of brands and contexts, we expect the models to learn the abstract concept of logoness and to be able to work with any logo class at inference time. In practice, we train these models in a class-agnostic way: every generated region proposal is classified in a binary fashion: logo or background discarding any class-related information.

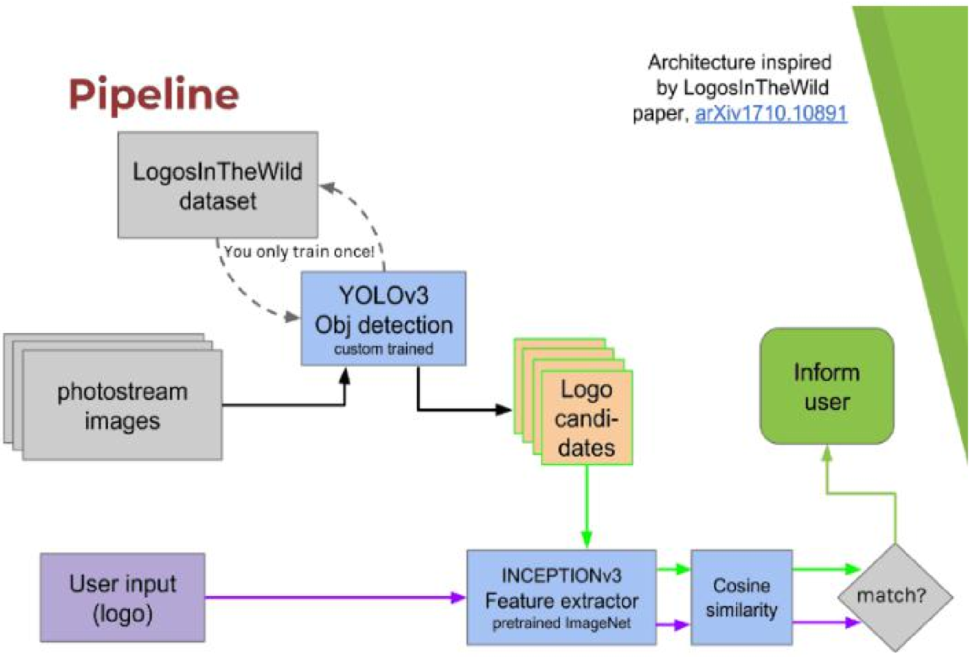
3.2 FEW SHOT LOGO RECOGNITION

Once the semantic logo detector has identified a set of probably logo regions within a given image we need to have a mechanism that can correctly classify these regions into its corresponding logo class/brand. Ideally, this step could be solved via a state of the art CNN image classification model such as ResNets[19] with multi-class classification. However this necessitates the right amount of training data for every class, class imbalance corrections and might also constrain the number of classes. Recent advances in deep embedding learning propelled the research in few- [30, 68] or one- [53] or zero-shot learning [33] where the aim is to use only a few, single or no examples of each class during training. The typical way this is achieved is via metric learning, where a model learns the similarity among arbitrary groups of data, thus being able to cope even with a large number of (unseen) classes. Currently, state-of-the-art methods for metric learning employ deep (convolution) neural networks, which are trained to output an embedding vector for each input image so that it minimizes a loss function defined over the distances of points. Usually, distances are learned using triplets of similar and dissimilar points D = (x, y, z), where x being the anchor, y the positive, and z the negative point and d is the Euclidean distance function. With y being more similar to x than z the task is to learn a distance respecting the similarity relationships encoded in D:

$$d(x, y) \leq d(x, z) \text{ for all } (x, y, z) \in D$$

Triplet-loss addresses this with a hinge function to create a fixed margin between the anchor-positive difference, and the anchor-negative difference:

$$L_{triplet}(x, y, z) = [d(x, y) + M - d(x, z)]_+$$



Chapter 4
Implementation

4.1 Logo Detection:

YOLOv3 was trained with randomly resized inputs in the range of [320, 640] with steps of 32, to achieve better accuracy. We computed the recall at IoU > 0.5, average precision (AP), and the number of regions generated on the negative/no logo set.

4.2 Few Shot Recognizer

For the few-shot logo embedding model we used the SEResnet50 achitecture with the same modifications . Input images were resized to 160x160 pixels, the embedding dimension was 128 and the batch size 32. We used the Adam optimizer with momentum 0.9, weight decay 0.0005, and learning rate 10−4 which we reduced by a factor of 0.8 every 20 epochs. The network's parameters were initialized using Xavier initialization with magnitude 2, no transfer learning was used. We trained few-shot model by passing our annotated images to the few-shot logo detector . The 28 logo classes were used for training and testing without any sampling strategy.

4.3 DataSet

This project uses the Logos In The Wild dataset which we requested via email directly from the authors of the paper, arXiv:1710.10891. This dataset includes 11,054 images with 32,850 bounding boxes for a total of 871 brands.

Chapter 5
Technologies

5.1 YOLO: Real-Time Object Detection

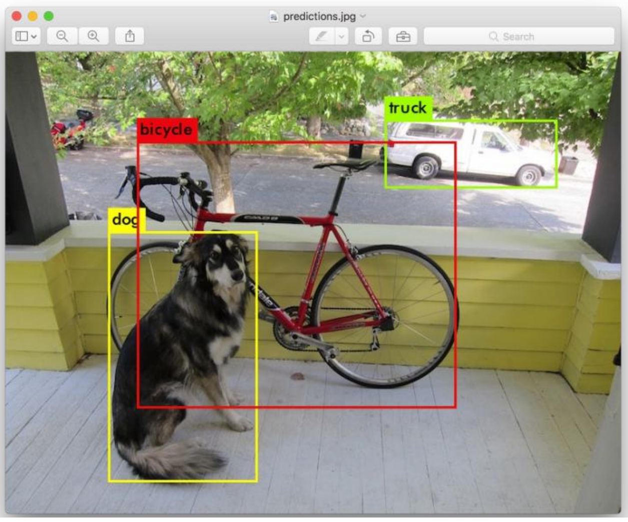
You only look once (YOLO) is a state-of-the-art, real-time object detection system. On a Pascal Titan X it processes images at 30 FPS and has a mAP of 57.9% on COCO test-dev.

YOLOv3 is extremely fast and accurate. In mAP measured at .5 IOU YOLOv3 is on par with Focal Loss but about 4x faster. Moreover, you can easily tradeoff between speed and accuracy simply by changing the size of the model, no retraining required!

Prior detection systems repurpose classifiers or localizers to perform detection. They apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections.

We use a totally different approach. We apply a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

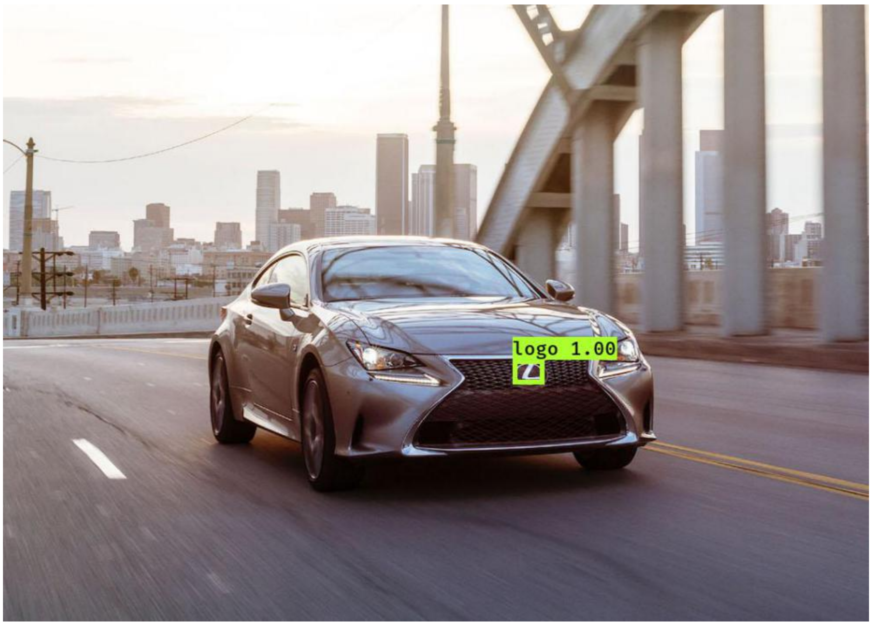
This model has several advantages over classifier-based systems. It looks at the whole image at test time so its predictions are informed by global context in the image. It also makes predictions with a single network evaluation unlike systems like R-CNN which require thousands for a single image. This makes it extremely fast, more than 1000x faster than R-CNN and 100x faster than Fast R-CNN. See our paper for more details on the full system.



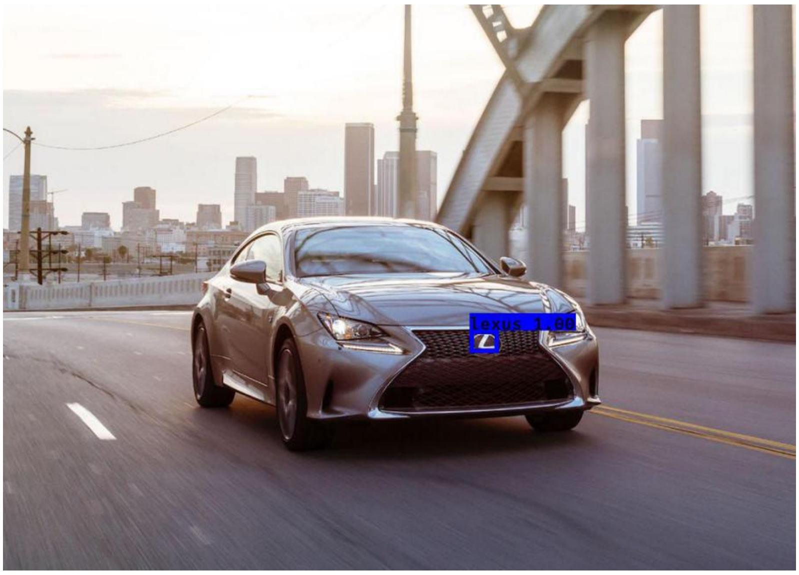
Chapter 6
Results and Recommendations

Test Image

Logo Detection



Logo recognition



Test Image

Logo Detected